## Artificial Intelligence

## and

## Machine Learning

Semester-IV (Batch-2022)

Case Study: train.mysql

<https://drive.google.com/file/d/1zwKpUEehTxX6h24D4eOZeQCaNLqrQCOX/view?usp=drive_link>

A red and white sign

Description automatically generated with low confidence

**By: Submitted By:**

Faculty Name Name : Simranpreet Kaur

Roll Number :2210990854

Group:G13(A)

**Department of Computer Science and Engineering**

## Chitkara University Institute of Engineering & Technology,

## Chitkara University, Punjab

**DESCRIPTION**

**Libraries:**

**Pandas:** Pandas is a Python library used for working with data sets. It has functions for analyzing, cleaning, exploring, and manipulating data.

The name "Pandas" has a reference to both "Panel Data", and "Python Data Analysis" and was created by Wes McKinney in 2008.

Pandas allows us to analyze big data and make conclusions based on statistical theories.

Pandas can clean messy data sets, and make them readable and relevant.

**Mysql.connector: mysql.connector** refers to the official MySQL Connector/Python, which is a Python driver provided by MySQL to connect and interact with MySQL databases. It enables Python applications to communicate with MySQL servers, allowing you to execute SQL queries, manage database connections, and handle data retrieval and manipulation.

**Questions:**

**1.**display top 5 rows of the database

2. check last 3 rows of the dataset

3.find shape of our dataset(number of rows & number of columns)

4 get information about our dataset like total number rows,total number of columns,datatype of each column and memory requiredment

5. get overall statistics about the dataframe

6. check null values in the dataset

7. drop the column

8. drop the column

9. how many people survived and how many died plot on graph

10. how many passengers were in first class,second class and third class plot those figures on graph

11. number of male and female passangers

12. bivariant analsis  
13. how has better cance of survival male or female

14. which passenger class has better chance of survival(first,second,or third class)?

**Methods:**

1.Sql.connect(): SQL.connect is a method that creates a connection to a SQL database using a connection string. A connection string is a set of parameters that specify how to connect to the database, such as the server name, database name, user name, password, and other options. SQL.connect returns a SqlConnection object that represents the established connection.

2. pandas.read\_sql(): **p**andas.read\_csv() is a function in the pandas library that reads a comma-separated values (csv) file into a Dataframe. It has many parameters that allow you to customize how the file is read, such as sep, header, index\_col, usecols, dtype, skiprows, nrows, na\_values, and more.

3.Data.head(): data.head(10) is a method that returns the first 10 rows of the DataFrame data. It is useful for getting a quick overview of the data or checking the format and column names. You can change the number of rows by passing a different value to the method, such as data.head(5) or data.head(20)

4.**data.tail(10) :** data.tail(10) is a method that returns the last 10 rows of the DataFrame data. It is useful for getting a quick overview of the data or checking the format and column names. You can change the number of rows by passing a different value to the method, such as data.tail(5) or data.tail(20)

**5.data.isnull().sum():** data.isnull().sum() is a method that returns the number of missing values in each column of the DataFrame data. It is useful for detecting and handling missing data in your analysis. It returns a Series with the column names and the count of null values.

6. Info() : the info() method is used to display a concise summary of a DataFrame, including the data types of columns, the number of non-null values, and memory usage.

7. describe() : describe() is a method in the pandas library of Python. It generates descriptive statistics for a DataFrame or a Series of numeric or object values. It summarizes the central tendency, dispersion, and shape of the data distribution, excluding NaN values. It also analyzes the unique, top, and frequency values for object data, such as strings or timestamps

8.Len(): The len() function is a built-in function in Python that returns the number of items in an object. It can be used with sequences, such as strings, lists, tuples, and ranges, or with collections, such as dictionaries, sets, and frozen sets.

9. unique() : unique() is a function that can be used to get the unique values from a list, a series, or an array in Python. It can be imported from different modules, such as pandas, numpy, or operator.

10. boxplot() : boxplot() is a function in the matplotlib.pyplot module of Python. It draws a box and whisker plot for a given data set. A box and whisker plot is a graphical representation of the distribution of a numerical variable, showing the median, quartiles, and outliers of the data

11. Groupby() : groupby() is used to group a pandas DataFrame or Series by one or more columns or levels and apply a function to each group.

12. Mean() : mean() is a function that can be used to calculate the arithmetic mean of a given set of numbers in Python. The arithmetic mean is the sum of the numbers divided by the number of elements

13.hist(): In Pandas, hist() is a method of the DataFrame class, and it is used to create histograms of the columns in the DataFrame.

14. pandas.get\_dummies():pandas.get\_dummies() is a function in the Pandas library for creating dummy/indicator variables from categorical variables. Dummy variables are binary columns that represent the presence or absence of a particular category in a categorical variable. This process is also known as one-hot encoding.

15.map(): In Pandas, map() is a method used to transform values in a Series using a dictionary, a function, or another Series. It is primarily used with Pandas Series to perform element-wise mapping and modification.

16.drop():The drop() method in Pandas is used to remove specified rows or columns from a DataFrame. It allows you to delete one or more specified rows or columns by specifying the labels or indices.